Calculation of Amino Acid pK_as in a Protein from a Continuum Electrostatic Model: Method and Sensitivity Analysis

P. BEROZA* and D. R. FREDKIN

Department of Physics, University of California, San Diego, La Jolla, California 92093

Received 30 June 1995; accepted 2 October 1995

ABSTRACT

We used continuum electrostatic theory to calculate pK_as of amino acids in protein. A Green's function formalism, based on a finite-difference solution to the Poisson–Boltzmann equation for a unit point charge, yields electrostatic potentials that allow calculation of amino acid pK_as to an estimated accuracy of tenths of a pK_a unit. Improvements over previous methods include the ability to focus the finite difference grid to arbitrarily small grid spacing, an analytical representation of the molecular surface, and a novel procedure to calculate the reaction field potential. Using this method, we performed a sensitivity analysis of calculated pK_as in the photosynthetic reaction center. Calculated pK_as are most sensitive for residues that are not well-exposed to solvent. Variations in the parameters of the continuum electrostatic model cause pK_a shifts that are larger than the accuracy of the numerical method, but probably not large enough to account for some of the discrepancies between calculated and experimentally measured pK_as that have been reported for the reaction center.

Introduction

Early models for calculation of the electrostatic potential in and around a solvated protein treated the protein as a spherical volume of low dielectric constant embedded in continuum of high dielectric constant.[1,2] Recently, refinements to this model have permitted inclusion of an atomic level of detail for both the position of charges within the protein and the shape of the dielectric interface.[3] The complex geometry of the protein–solvent interface (i.e., the dielectric boundary) prohibits analytical calculation of the electrostatic potential, which can be determined numerically using a number of methods (e.g., finite-difference,[3]

*Author to whom all correspondence should be addressed at the Department of Molecular Biology, Scripps Research Institute, La Jolla, CA 92037.

CCC 0192-8651 / 96 / 101229-16

finite-element,[4] and boundary-element[5]).

Such models have been applied to calculate the pK_as of individual amino acid residues in a protein.[6-11] Titration of an amino acid alters the net charge on the amino acid by one electronic charge (e). Therefore, it seems likely that the difference in the pK_a of a particular amino acid residue in a protein from its value in solution results from electrostatic interactions. Predicting the size and physical basis for these pK_a shifts is a particularly challenging problem for continuum models because it involves electrostatic fields *inside* the protein. However, little effort has been made to assess the accuracy, convergence, and sensitivity of the computational method as applied to the calculation of pK_as.

In this article we present a Green's function formalism for the problem of amino acid titration in a protein. A Green's function approach to this problem has been presented previously,[8] and we focus here on how this formalism allows the electrostatic energy to be separated into pieces that can be calculated with high accuracy and for which numerical convergence can be demonstrated. We show, for spherical test systems, that the calculated electrostatic potentials agree well with analytical solutions. Once the accuracy of the numerical method has been established, we examine the sensitivity of calculated pK_a values to small variations in the parameters of the continuum electrostatic model. The issues addressed came about from our efforts to understand the coupling of electron transfer with protonation changes of amino acid residues in the photosynthetic reaction center (RC) of *Rb. sphaeroides* (see ref. 12 for a full discussion). The RC is an interesting system for amino acid titration because it has a large number (> 100) of titratable residues, many of which are not well-exposed to solvent.

Continuum Electrostatic Model of Protein Titration

The average protonation of an amino acid, which is an equilibrium property, can be expressed as a statistical average over the accessible states of the protein:

$$\langle x_i \rangle = \frac{\sum_{\xi} x_i(\xi) e^{-G(\xi)/k_B T}}{Z} \qquad (1)$$

where $\langle x_i \rangle$ is the average protonation state of titrating site i, $x_i(\xi)$ is the protonation state of site i in protein state ξ, $G(\xi)$ is the free energy of this state, and the sum is over all such states. The distribution is normalized by the partition function: $Z = \sum_{\xi} e^{-G(\xi)/k_B T}$, where k_B is Boltzmann's constant and T is the temperature.

In defining the accessible states of the protein we make four major assumptions:

1. The electrostatic energies of the different protonation states of the protein are not affected by protein motion and may be calculated from the static X-ray structure.
2. The state of a protein is uniquely defined by the protonation state of each titratable residue.
3. The change in pK_a of an amino acid when it is bound to a protein is entirely electrostatic in origin.
4. The dielectric response of the protein and solvent is approximately that of a dielectric continuum, and the distribution of counterions is given by the Debye–Hückel theory.

Given these restrictions, the state of a protein can be defined by a vector **x** whose elements specify the number of titratable protons bound to each titrating site (either 0 or 1). For a protein with N titrating sites, we have:

$$\mathbf{x} = (x_1, x_2, \ldots, x_N) \qquad (1)$$

where x_i is the protonation state of site i. Each element has two values; therefore, 2^N terms enter eq. (1). Evaluation of eq. (1) at a number of pH values yields a titration curve.[6] However, for all but the smallest proteins, computing the sum in eq. (1) presents a problem, because the number of terms in the sum grows exponentially with the number of titrating sites. Therefore, computing the complete sum is only feasible for small proteins (e.g., lysozyme[6]). For larger proteins, approximation methods must be employed. The mean-field approximation (also known as the Tanford–Roxby approximation[13]) has been shown to fail for strongly coupled sites.[9,14] Methods that either discard some of the unlikely states in the partition function[14] or treat clusters of interacting sites exactly, while using the mean-field approximation for interactions between sites in different clusters,[9,15] succeed in calculating the sum for most proteins. However, these methods fail for proteins that have large clusters of strongly interacting sites (e.g., the photosynthetic reaction center). We have

demonstrated that Monte Carlo sampling of protonation states gives reliable titration curves for very large proteins with strongly interacting sites,[7] and we employ this method in the current work.

Green's Function Formalism for $G(\mathbf{x})$

To compute the average protonation from eq. (1), one needs to be able calculate the electrostatic energy of each protonation state of the protein. As described above, the continuum electrostatic model assumes that, at fixed pH, this energy is a function of the protonation state ($\mathbf{x}$) of the protein [i.e., $G(\xi) = G(\mathbf{x})$]. For a given protonation state $\mathbf{x}$ the charge distribution in the protein is fixed, and the electrostatic energy is determined by electrostatic potential at these charges, which in the continuum model is determined by: (i) the dielectric constant; and (ii) the distribution of free charge, $\rho(\mathbf{r})$. In addition, numerical calculations are often confined to a finite region, and then the potential is also determined by the boundary conditions on the surface enclosing the region.

In the theory of electrostatics of continuous media, the free energy of a charge distribution is given by:

$$G_{es} = \frac{1}{2}\int\int \rho(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho(\mathbf{r}') d^3r d^3r' \qquad (2)$$

where $\rho(\mathbf{r})$ is the charge density at position $\mathbf{r}$ and $g(\mathbf{r},\mathbf{r}')$ is the electrostatic Green's function (see, e.g., ref. 16). For a protein, we separate the charge distribution into two pieces:

$$\rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \sum_i Q_i \rho_i(\mathbf{r}) \qquad (3)$$

where $\rho_0(\mathbf{r})$ is the charge distribution on the protein when all the titrating sites are neutral, $Q_i \rho_i(\mathbf{r})$ is the correction to the charge distribution when titratable site i is charged, and Q_i is the net charge on the site (-1, 0, or 1). In our computations we assume that $\rho_i(\mathbf{r})$ is confined to the titrating site. Substituting eq. (3) into (2):

$$G_{es} = \int\int \left\{ \frac{1}{2}\rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_0(\mathbf{r}') + \sum_i \rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') Q_i + \frac{1}{2}\sum_{i,j} \rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_j(\mathbf{r}') Q_i Q_j \right\} d^3r d^3r'$$

Separation of $\rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_j(\mathbf{r}')$ into diagonal and off-diagonal terms gives:

$$G_{es} = \int\int \left[\frac{1}{2}\rho_0(r) g(\mathbf{r},\mathbf{r}') \rho_0(\mathbf{r}') + \sum_i Q_i \left\{ \rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}) + \frac{1}{2}\sum_{i\in C} \rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') - \frac{1}{2}\sum_{i\in A} \rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') \right\} + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}} \rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_j(\mathbf{r}') Q_i Q_j \right] d^3r d^3r' \qquad (4)$$

where C and A represent the set of cationic (i.e., positively charged when protonated) and anionic (i.e., neutral when protonated) sites, respectively. To write this in terms of the protonation vector $\mathbf{x}$ we substitute $Q_i = x_i$ for cationic sites and $Q_i = x_i - 1$ for anionic sites. We define the intrinsic proton binding energy of a site as:

$$\epsilon_i = \int\int \left[\rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') \pm \frac{1}{2}\rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') \right] d^3r d^3r' \qquad (5)$$

(+ for cations, − for anions) and the interaction energy between a pair of charged sites to be:

$$W_{i,j} = \int\int \rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_j(\mathbf{r}') d^3r d^3r' \qquad (6)$$

Substitution of eqs. (5) and (6) into eq. (4) gives:

$$G_{es} = \frac{1}{2}\int\int \rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_0(\mathbf{r}') d^3r d^3r' - \sum_{i\in A} \epsilon_i + \sum_i x_i \epsilon_i + \frac{1}{2}\sum_{i,j,i\neq j} W_{ij} Q_i Q_j \qquad (7)$$

This equation represents only the total electrostatic energy of the protein and does not account for the chemical contributions of the individual titrating sites to the proton binding energy. We incorporate these contributions into the intrinsic energy term (ϵ_i) by using the isolated amino acid in solution as a reference state. We assume that the free energy of proton binding can be separated

into electrostatic and nonelectrostatic pieces. If we assume that the nonelectrostatic contribution to the proton binding affinity of an amino acid is independent of its environment (i.e., bound to a protein or isolated in solution), then we can obtain the proton binding free energy in the protein from:

$$\Delta G_p^{A \to AH^+} = \Delta G_s^{A \to AH^+} + \epsilon_{i,p} - \epsilon_{i,m} \quad (8)$$

where $\Delta G_p^{A \to AH^+}$ is the intrinsic proton binding energy of titrating site i in the protein; $\Delta G_s^{A \to AH^+}$ is the proton building energy of the titrating amino acid in solution [i.e., $\Delta G_s^{A \to AH^+} = -k_B T(\ln 10)\, pK_{a,\text{soln}}$, where $pK_{a,\text{soln}}$ is the experimentally measured pK_a of the isolated amino acid in solution]; $\epsilon_{i,p}$ is the value of eq. (5) in the protein; and $\epsilon_{i,m}$ is value of eq. (5) in the "model compound" (i.e., the isolated amino acid in solution[6]).

Equation (8) is related to the intrinsic pK_a, pK_{int}, defined by Tanford as the pK_a a site would have when all other titratable sites are in their neutral state, by:

$$pK_{\text{int}} = pK_{a,\text{soln}} - \frac{1}{k_B T(\ln 10)}(\epsilon_{i,p} - \epsilon_{i,s}) \quad (9)$$

When the chemical potential of the protons in solution is included, we have, for the free energy of a protonation state $\mathbf{x}$:

$$G(\mathbf{x}) = G_0 + \sum_i x_i(-k_B T \ln 10)(pK_{\text{int},i} - \text{pH}) + \frac{1}{2}\sum_{i,j}(x_i + q_i^u)(x_j + q_j^u)W_{i,j} \quad (10)$$

where the constant terms in eq. (7) have been combined as G_0, x_i is the protonation state of site i in protein state $\mathbf{x}$, $pK_{\text{int},i}$ is the intrinsic pK_a of site i, q_i^u is the charge on site i in its unprotonated state, and $W_{i,j}$ is the interaction energy between two charged sites. This furnishes a separation of the electrostatic energy of protonation state $\mathbf{x}$ into two pieces: (i) the intrinsic pK_a, which, for each amino acid, is independent of the protonation state of other amino acids; and (ii) site–site interactions (W_{ij}), which depend on which amino acids are charged.

The intrinsic pK_as and site–site interactions are calculated from continuum electrostatic theory by evaluating the integrals [eq. (5)] for both the protein and model compound and the integrals [eq. (6)] for the protein. When all charges and dipoles are represented as point charges located at the atomic positions (obtained from the X-ray structure), the charge distribution, $\rho(\mathbf{r})$, reduces to a sum of δ-functions. The charge distribution of the protein when all residues are neutral becomes:

$$\rho_0(\mathbf{r}) = \sum_\alpha q^0(\alpha)\delta(\mathbf{r} - \mathbf{r}_\alpha) \quad (11)$$

where the sum is over all atoms in the protein, and q^0 is the charge on the atom in the neutral state (i.e., no titratable residues ionized). The piece of the charge distribution that depends upon the protonation state ($\mathbf{x}$) of the protein becomes:

$$Q_i \rho_i(\mathbf{r}) = Q_i \sum_{\alpha \in i} \Delta q_i(\alpha)\delta(\mathbf{r} - \mathbf{r}_\alpha) \quad (12)$$

where the sum is over all atoms on the charged residue i. The term $\Delta q_i(\alpha)$ is the difference between the partial charge of an atom in the protonated and unprotonated states of the amino acid; i.e., $\Delta q_i(\alpha) = [q_i^p(\alpha) - q_i^u(\alpha)]$, where $q_i^p(\alpha)$ and $q_i^u(\alpha)$ are the partial charges on the atom when the residue is protonated or unprotonated.

The δ-functions in the integrands in eqs. (5) and (6) reduce the integrals to pairwise electrostatic interactions between atoms. The intrinsic energy term becomes:

$$\begin{aligned}
\epsilon_i &= \int\int \rho_0(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') \\
&\quad \pm \frac{1}{2}\rho_i(\mathbf{r}) g(\mathbf{r},\mathbf{r}') \rho_i(\mathbf{r}') d^3r d^3r' \\
&= \int\int \Bigg\{ \sum_\alpha q^0(\alpha)\delta(\mathbf{r} - \mathbf{r}_\alpha) g(\mathbf{r},\mathbf{r}') \\
&\quad \times \sum_{\beta \in i} \Delta q_i(\beta)\delta(\mathbf{r} - \mathbf{r}_\beta) \qquad (13) \\
&\quad \pm \frac{1}{2}\sum_{\alpha \in i}\sum_{\beta \in i} \Delta q_i(\alpha)\Delta q_i(\beta)\delta(\mathbf{r} - \mathbf{r}_\alpha) \\
&\quad \times g(\mathbf{r},\mathbf{r}')\delta(\mathbf{r}' - \mathbf{r}_\beta) \Bigg\} d^3r d^3r' \\
&= \sum_{\beta \in i} \Delta q_i(\beta) \Bigg\{ \sum_\alpha q^0(\alpha)\Phi(\alpha,\beta) \\
&\quad \pm \frac{1}{2}\sum_{\alpha \in i} \Delta q_i(\alpha)\Phi(\alpha,\beta) \Bigg\} \qquad (14)
\end{aligned}$$

where the sums are over all atoms in the protein (or model compound) unless specified. In the last line of the equation we have replaced the notation $g(\mathbf{r}_\alpha, \mathbf{r}_\beta)$ with $\Phi(\alpha,\beta)$ to emphasize that it is the electrostatic potential at atom α resulting from a unit charge at atom β [note that $\Phi(\alpha,\beta) = \Phi(\beta,\alpha)$]. We redefine $\Phi(\alpha,\alpha)$, which is singular,

to be the reaction field potential at atom α (i.e., the potential at the atom resulting from the polarization it induces in the dielectric media and the screening charge it induces in the solvent). The magnitude of the interaction between sites is given by:

$$W_{i,j} = \sum_{\alpha \in i} \sum_{\beta \in j} \Delta q_i(\alpha) \Delta q_j(\beta) \Phi(\alpha, \beta) \quad (15)$$

Thus, the values $pK_{\text{int},i}$ and W_{ij} can be expressed in terms of Green's functions. The Green's function, i.e., the electrostatic potential at $\mathbf{r}$ generated by a unit charge at $\mathbf{r}'$, is defined by the Poisson–Boltzmann equation with a unit source charge:

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla g(\mathbf{r}, \mathbf{r}') - \epsilon(\mathbf{r}) \kappa(\mathbf{r})^2 g(\mathbf{r}, \mathbf{r}') = -4\pi\delta(\mathbf{r} - \mathbf{r}') \quad (16)$$

When the charge distribution of an amino acid is represented by a set of atom-centered point charges, we need only concern ourselves with the Green's function between points $\mathbf{r}$ and $\mathbf{r}'$ that coincide with the atom centers. Further, Green's functions between atoms whose charge is constant, regardless of the protonation state, may be ignored because they add an overall constant to the energy of any state. Thus, the Green's function, $g(\mathbf{r}, \mathbf{r}')$, is evaluated at all points $\mathbf{r}$ that coincide with atom centers and points $\mathbf{r}_\alpha$ that coincide with atoms in a titrating residue. The number of Green's functions calculated equals the number of atoms in titrating residues in the protein whose partial charge changes when the residue titrates. Each Green's function is calculated by placing a unit charge on the titrating atom and computing the potential it generates at all the atoms in the protein.

The Green's functions are used to determine the intrinsic pK_as and the matrix of site–site interactions (i.e., W_{ij}). Once the intrinsic pK_as and W_{ij} terms are obtained from the Green's functions $\Phi(\alpha, \beta)$, the energy of a protonation state can be calculated from eq. (10).

Finite Difference Method

The geometry of the molecular surface is too complicated for analytical solution of the Poisson–Boltzmann equation; numerical methods must be used. We have chosen the finite difference method for solving the Poisson–Boltzmann equation, which has been successfully used to determine electrostatic potentials in biomolecules[3,17–19] and to calculate pK_as of amino acids in proteins.[6–9]

In the finite difference method, space is discretized on a three-dimensional, uniformly spaced grid. Each grid point lies at the center of a cubical volume, whose edges have length equal to the grid spacing (h). A grid point (with indices i, j, k) and its neighboring grid points are shown in Figure 1.

The finite difference equations are obtained by integrating the linear Poisson–Boltzmann equation within the volume (V) associated with a grid point (i.e., the set of all points in space that are closer to a grid point than to any other grid point):

$$\int_V [\nabla \cdot \epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r})]\, dV - \int_V [\epsilon(\mathbf{r}) \kappa(\mathbf{r})^2 \Phi(\mathbf{r})]\, dV = \int_V [-4\pi\rho(\mathbf{r})]\, dV \quad (17)$$

Application of the divergence theorem to the first integral yields a surface integral over the faces of the cube, $\int_S [\epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r}) \cdot \hat{n}]\, dA$, where $\hat{n}$ is the outward-pointing normal from the surface. The second integral vanishes for grid points that do not lie within the volume accessible to counterions; otherwise, it is approximated by $h^3 \epsilon_w \kappa^2 \Phi(i, j, k)$, where κ is the Debye–Hückel screening constant ($\kappa^2 = 8\pi e^2 N_A I / \epsilon_w k_B T$, where e is the electronic charge, N_A is Avogadro's number, I is the ionic strength, ϵ_w is the dielectric constant of water, k_B is the

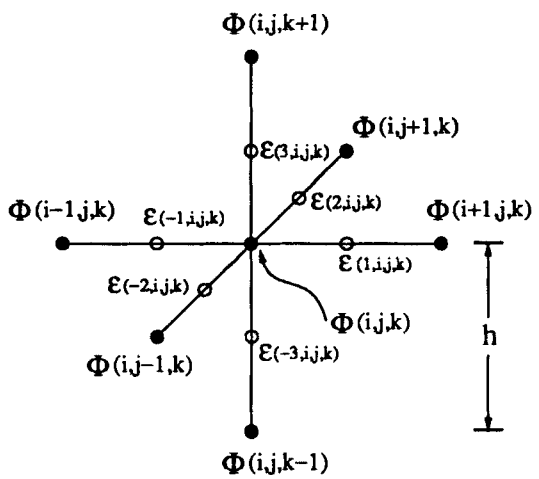

FIGURE 1. A finite difference grid point with its nearest neighbors (solid circles). The potential $\Phi(i, j, k)$ is related to the potential at its neighbors by eq. (18), which results in a set of coupled equations that are solved numerically. The midpoints between Φ grid points make up the "dielectric grid" (open circles); the values of ϵ are assigned based on each midpoint's solvent accessibility. Grids for the salt concentration and the charge distribution coincide with the Φ grid.

Boltzmann constant, and T is the temperature). The third integral is approximated by $-4\pi q(i, j, k)$, where $q(i, j, k)$ is the total charge enclosed with the volume. Finite difference methods in which atom-centered charges are distributed on the eight nearest grid points by a trilinear function[20] can lead to situations in which charge density associated with atoms within the protein lies *outside* the protein dielectric, which is not permitted by the model. With these approximations eq. (17) becomes:

$$\sum_{6 \text{ faces}} \int_S [\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) \cdot \hat{n}]\, dA - \epsilon_w \kappa(i, j, k)^2 \Phi(i, j, k) = -\frac{4\pi q(i, j, k)}{h^3}$$

where $\kappa(i, j, k)$ is the value of the Debye–Hückel screening constant at the grid point (i, j, k). The gradient is approximated by a finite difference between the potential at (i, j, k) and the potential at the neighboring point on the other side of the face. For example, the surface integral on the face between grid points (i, j, k) and $(i + 1, j, k)$ is approximated by:

$$\int_S [\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) \cdot \hat{n}]\, dA = \epsilon(1, i, j, k)\frac{[\Phi(i + 1, j, k) - \Phi(i, j, k)]}{h} h^2$$

where $\Phi(i, j, k)$ is the value of the potential at grid point (i, j, k) and $\epsilon(1, i, j, k)$ is the value of the dielectric constant at the center of the face [i.e., the midpoint of the line joining grid points (i, j, k) and $(i + 1, j, k)$ in Fig. 1]. Similar equations are obtained for the other five faces. Equation (17) becomes a linear algebraic equation:

$$\begin{aligned}
&\{\epsilon(1, i, j, k)[\Phi(i + 1, j, k) - \Phi(i, j, k)] \\
&\quad -\epsilon(-1, i, j, k)[\Phi(i, j, k) - \Phi(i - 1, j, k)]\}/h^2 \\
&\quad +\{\epsilon(2, i, j, k))[\Phi(i, j + 1, k) - \Phi(i, j, k)] \\
&\quad -\epsilon(-2, i, j, k)[\Phi(i, j, k) - \Phi(i, j - 1, k)]\}/h^2 \\
&\quad +\{\epsilon(3, i, j, k))[\Phi(i, j, k + 1) - \Phi(i, j, k)] \\
&\quad -\epsilon(-3, i, j, k)[\Phi(i, j, k) - \Phi(i, j, k - 1)]\}/h^2 \\
&\quad -\epsilon_w \kappa(i, j, k)^2 \Phi(i, j, k) = -4\pi q(i, j, k)/h^3
\end{aligned} \tag{18}$$

Using this notation, the dielectric grid point $\epsilon(-1, i, j, k)$ is the same as $\epsilon(1, i - 1, j, k)$.

The potentials at the grid points are determined by the set of coupled linear algebraic equations [eq. (18)]. Determination of the electrostatic potential at each grid point for a cubic grid with N grid points in each dimension requires three arrays of size N^3 to store values for Φ, κ, and q and an array of size $\sim 3N^3$ to store the values of the dielectric constant between neighboring grid points. We label the grid of dimension N the "potential grid" and the grid of dimension $3N$ the "dielectric grid."

For a potential grid of dimension N there are N^3 coupled equations which we solved using a diagonally scaled conjugate gradient solver.* We considered the solution to have converged when the 2-norm of the residual divided by the 2-norm of the right-hand side of (18) was $< 10^{-5}$. After the potential was determined for each point on the finite difference grid, the potential at each atom in the protein was found by linear interpolation of the potentials at the eight nearest grid points.

We computed the Green's function for each titrating atom (i.e., an atom whose charge changes upon titration of the residue), which is equivalent to calculating the electrostatic potential at all atoms in the protein resulting from a unit charge on the titrating atom. For each Green's function, we center the grid on the source atom. Therefore, the only grid point that has charge is the origin. The Debye–Hückel treatment of the salt is contained in the $\kappa(i, j, k)$ term which is 0 if the grid point lies within the ion-exclusion layer, which we take to be 2 Å beyond the molecular surface of the molecule.[19] The dielectric constant at the midpoint between two grid points is set to ϵ_w if that point is accessible to solvent, otherwise it is set to ϵ_p, the dielectric constant of the protein.

BOUNDARY CONDITIONS

The boundary conditions are the remaining specifications for a unique electrostatics problem.[16] Two boundaries are involved in the calculation of the electrostatic potential in a protein: (i) the protein–solvent interface and (ii) the boundary at in-

*FORTRAN subroutines were obtained from SLAP (Sparse Linear Algebra Package), version 2.0, written by M. K. Seager (Lawrence Livermore National Laboratory) and A. Greenbaum (NYU).

finity. The proper boundary conditions at the protein–solvent interface are included in the derivation of the finite-difference equations, while the boundary condition at infinity (that the potential vanish) must be approximated because only finite volumes can be handled numerically with the finite-difference method. Screening by counterions guarantees that, in the solvent, the potential falls off exponentially with distance from the protein surface. For a large enough volume, the potential at the boundary of the volume will be essentially zero. We choose the volume large enough such that its surface is at least 2 Debye screening lengths from the surface of the protein, and, therefore, from any source charge. For this choice of grid dimension we found no significant difference between setting the boundary potential to zero or setting it to the screened coulomb potential $(e^{-\kappa r}/\epsilon_w r)$.

CALCULATION OF DIELECTRIC GRID

One of the difficulties that arises when solving the Poisson–Boltzmann equation for a complicated macromolecule like a protein is the determination of the extent of the low dielectric volume of the protein. The region internal to the protein is the volume from which solvent molecules are excluded by the protein atoms because of the Pauli exclusion principle. The coordinates of the protein atoms provide a starting point for determination of the interface between the two dielectric regions. The degree to which solvent can penetrate the protein is determined by modeling protein atoms and solvent molecules as hard spheres. The excluded volume for each protein atom is given by its van der Waals radius. The solvent is represented by a sphere of radius 1.4 Å, and any point that can be touched by this spherical probe without causing van der Waals overlap between the probe and any protein atoms is considered to be accessible to solvent and is assigned a high dielectric constant. Thus, the dielectric interface between protein and solvent is equivalent to the molecular surface, as defined by Richards.[21] Note that, in addition to the "external surface," there may be cavities within the protein which are assigned to the high dielectric region.

Each dielectric grid point lies internal or external to the molecular surface of the protein and is assigned the protein or solvent dielectric constant accordingly. Points that lie within the van der Waals radius of a protein atom are definitely inaccessible, and those that lie beyond the sum of the solvent radius and the van der Waals radius of the closest protein atom are definitely accessible. Determination of the dielectric constant for the remaining grid points (those that lie within the probe radius of the van der Waals surface of the protein) is a complicated geometrical problem.

Our method for solving this problem relies on the work of Connolly to determine an analytical expression for the molecular surface of a molecule.[22] The molecular surface can be separated into three parts, depending on how many atoms the solvent probe is in contact with when it touches a point on the molecular surface. If the solvent probe is in contact with only one atom, the molecular surface is convex and coincides with the van der Waals surface of the atom. If the probe contacts two atoms simultaneously it touches an arc on a toroidal section of the molecular surface (this torus is swept out as the probe sphere "rolls" maintaining contact with both atoms). Finally, if the probe is in contact with three atoms simultaneously, it defines a concave surface between the three atoms. Methods for calculating these convex, toroidal, and concave sections of the molecular surface are described by Connolly,[22] and we apply them here to determine the geometric relationship between the dielectric grid points and the molecular surface.

First we determine the set of points that correspond to the center of the solvent probe when it is in simultaneous contact with three protein + atoms with no van der Waals overlap with other protein atoms (the details of how these points are determined can be found in Connolly's article[22]). To help keep the geometrical constructs clear, we call these points "three-contact" points, because the solvent probe touches three protein atoms when its center is at such a point.

Second, we determine the toroidal sections of the solvent accessible surface. The solvent probe can make simultaneous contact with two atoms, if they are close enough, and trace out a torus as it rolls around the two atoms. This torus is centered on the axis joining the two atomic centers. Its geometry depends upon the van der Waals radii of the two atoms, the distance between them, and the solvent probe radius. All, some or none of the volume in the torus may be accessible, depending on the position of nearby atoms and their overlap with the solvent probe as it traces out the torus. For tori that have partial accessibility, the arc that the center of the solvent probe traces out as it rolls through the toroidal volume begins and ends at two "three-contact" points described above. Thus,

we have a set of circular arcs that join "three-contact" points (additionally, there are a few free-standing tori, which have no intersection with the van der Waals radius of any atom; these arcs form complete circles).

These geometric entities need only be calculated once, for they are independent of the finite difference grid and the level of detail (i.e., focusing) of the problem. They are stored in computer memory to help assign dielectric constants to the dielectric grid points.

The rules for assigning dielectric constants to dielectric grid points are:

1. Grid points within the van der Waals radius of a protein atom are internal.
2. Grid points that lie beyond the sum of the solvent radius and the van der Waals radius of the closest protein atom are external.
3. Grid points that lie within the solvent probe radius of a "three-contact point" are external.
4. Grid points within solvent-accessible volumes of tori are external (see below).
5. Grid points that lie within sum of the solvent probe radius and the van der Waals radius of only one protein atom are external (these are points closest to the convex part of the molecular surface).
6. All other points are internal.

Several techniques help speed up the above procedure. First, only rules 1 and 2 are applied, and the points left unresolved by these two rules are marked as uncertain. For these uncertain points, the remaining rules are applied in order (each successive rule is more time-consuming). In addition, neighbor lists are constructed for each atom, and, for the undetermined grid points, only tori and "three-contact points" that make contact with the atom nearest to the grid point or its neighbors are considered in the rules above.

Rule 4 merits further comment, as it is the most geometrically complicated. For each torus, part of its volume may be accessible and part inaccessible. If a dielectric grid point is within the volume of the torus, its status (either in the low dielectric region or in the external region) can only be determined if it is within an *accessible* volume of the torus. If it is not, it may still be an accessible point, and, therefore, its status remains unresolved (e.g., a point that is within the inaccessible volume of one torus may be within the accessible volume of a different torus). A grid point is within the accessible volume of the torus if it is within a solvent probe radius of an accessible arc that the solvent probe center traces out as it sweeps through the toroidal volume.

Although this method requires some complicated geometrical calculations of the molecular surface of the protein, they need be performed only once, and the result can be applied to any finite difference grid to determine the dielectric constant. No approximations are involved in assigning dielectric constants, because each dielectric grid point is compared to the exact analytical surface that defines the dielectric interface.

FOCUSING

There are three conflicting requirements of the finite difference grid: (1) its spacing must be coarse enough to include the entire protein; (2) its spacing must be fine enough to give accurate potentials near the source charge; and (3) its size must be small enough to permit storage of its values in the computer's random access memory.

To resolve these conflicts we apply the method of focusing.[23] The initial grid, which is small enough to store in the computer memory, is given a spacing large enough to insure that no part of the protein surface is within 2 Debye screening lengths of the grid boundary. Once the potentials are found for this initial grid, we use a finer grid, which has the same number of points as the coarse grid and is also centered on the source charge (in practice, these grids occupy the same area in the computer memory). We set the boundary potentials on the fine grid by linear interpolation from the coarse grid. Once the boundary potential has been set for the finer grid, we have a new electrostatics problem. The arrays for the salt, charge, and dielectric constants are recomputed for the new grid. The source charge still lies at the center of the grid, but more grid points now lie within the low dielectric region, closer to the charge. The new coupled set of equations is solved by the methods previously discussed. A two-dimensional example of focusing is shown in Figure 2. The focusing procedure can be repeated, yielding more accurate values for the potential near the source charge. The most accurate potential at each atom is determined from the finest grid that contains the atom within its volume.

For choice of a focusing factor (i.e., the reduction factor for the grid spacing in successive grid refinements) we consider the accuracy of the finite difference solution for the potential from a point

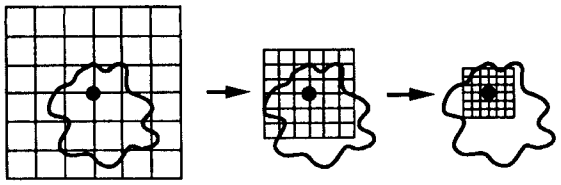

FIGURE 2. Example of focusing. Shown are three levels of focusing of the finite difference grid. The grids are centered on the point charge for which the Green's function is being calculated. The boundary of each refined grid is initialized from the potential on the previous coarse grid.

charge in a homogeneous dielectric medium. Although the grid potentials begin to deviate from the true potential ($\frac{1}{\epsilon r}$) within half the grid extent, we have found that a focusing factor of $\frac{1}{4}$ does not lead to significant errors in the calculated amino acid pK_as.

BOUNDARY INTEGRAL METHOD

The source charge interacts with its reaction field (created by the polarization it induces at interfaces between regions of differing dielectric constant and by the distribution of counterions it causes in solution), which stabilizes the source charge. The resulting electric field is called the reaction field,[24] and the potential at the source atom is the reaction field potential.

In finite difference solutions for the electrostatic potential, the reaction field potential is combined with the potential from the source charge, the Coulomb potential. At grid points away from the charge, the Coulomb potential is known and may be subtracted off to obtain the reaction field potential at that point. However, when $\alpha = \beta$, the $\Phi(\alpha, \beta)$ term in eq. (13) is the reaction field potential at the source charge, where the Coulomb potential diverges. It is not possible, therefore, to subtract the Coulombic term from the total potential. In fact, the total potential at the source charge is finite only by virtue of the discrete grid. In the limit of small grid spacing ($h \to 0$), the total potential also diverges.

How does one separate the reaction field potential from the diverging Coulomb potential? A common technique is to perform a separate finite difference calculation with only the protein dielectric present (and no dielectric interface).[25] The grid potential at the charge represents the diverging Coulomb potential (actually, this potential can be calculated more quickly using other numerical techniques[26]). This is then subtracted from the combined potential calculated with the dielectric interface present, leaving only the reaction field potential at the source charge.

Alternatively, we can find the reaction field potential in a single set of finite difference focusing iterations. The reaction field potential can be obtained from a boundary integral calculated on a surface enclosing the charge. We start with Green's theorem (see Jackson[16]):

$$\int_V (\phi\nabla^2\psi - \psi\nabla^2\phi)\, d^3x = \oint_S \left[\phi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\phi}{\partial n}\right] da \tag{19}$$

where ϕ and ψ are arbitrary scalar fields in volume V enclosed by surface S, and $\frac{\partial}{\partial n}$ is the outward normal derivative on the surface S. If we let $\psi = \frac{1}{\epsilon R}$, the Coulomb potential from the point charge (where ϵ is the dielectric constant of the medium and R is the distance from the source charge, at $\mathbf{x}_q$, to the point where the potential is being calculated), and $\phi = \Phi_{RF}$, the reaction field potential, and we choose the volume V to be small enough for ϵ to be a constant within it. The left-hand side of eq. (19) becomes:

$$\frac{1}{\epsilon}\int_V \left(\Phi_{RF}\nabla^2\frac{1}{R} - \frac{1}{R}\nabla^2\Phi_{RF}\right) d^3x = \frac{1}{\epsilon}\int_V - 4\pi\delta(\mathbf{x} - \mathbf{x}_q)\Phi_{RF}\, d^3x = \frac{-4\pi}{\epsilon}\Phi_{RF}(\mathbf{x}_q) \tag{20}$$

The Laplacian of $1/R$ produces a δ-function; and the Laplacian of Φ_{RF} is 0 because Φ_{RF} satisfies Laplace's equation in volume V.

Substitution of eq. (20) into eq. (19) gives an expression for the reaction field potential at the source charge in terms of an integral of the reaction field potential and its normal derivative on any surface enclosing the charge:

$$\Phi_{RF}(\mathbf{x}_q) = \frac{1}{4\pi}\oint_S \left[\frac{1}{R}\frac{\partial}{\partial n}\Phi_{RF} - \Phi_{RF}\frac{\partial}{\partial n}\frac{1}{R}\right] da \tag{21}$$

The dielectric constant ϵ divides out because ϵ is constant in volume V. On the surface of integration, $\Phi_{RF} = \Phi_{FD} - \Phi_0$, where Φ_{FD} is the potential given by the finite difference calculation and Φ_0 is the Coulomb potential of the point charge. Since $\Phi_0 = \frac{1}{\epsilon R}$, its contribution is zero and Φ_{RF} can be replaced by Φ_{FD} in eq. (21). Thus, the reaction field potential at the point charge can be found by

evaluating the above integral over a surface enclosing the charge.

A cube, chosen to coincide with the faces of the finite difference grid, provides a convenient surface of integration because the potential and its derivatives are easily computed from the finite difference grid. The finite difference calculation is most accurate far from the charge, so we want to evaluate eq. (21) on the faces of the largest subgrid in the finite difference grid that lies entirely within the protein. Because the initial grid must be coarse to ensure accurate boundary conditions, several focusing iterations are required to obtain a subgrid fine enough to evaluate eq. (21).

Testing the Finite Difference Method

Analytical solutions to the Poisson–Boltzmann equation exist for spherically symmetric systems,[2] and we use these to test the accuracy of the finite-difference technique. Following the example of Gilson et al.,[23] we represent the protein as a low dielectric sphere of radius 30 Å. Although the atomic detail of the molecular surface of the protein and amino acid is lost, the system provides a way of estimating the error in the finite-difference method, which, one hopes, is approximately the same when the dielectric interface is too complicated to permit analytical evaluation of the electrostatic potential. The question is: How accurate is the finite difference method in determining the intrinsic binding energies, i.e., the ϵ_i terms in (7) and the site–site interactions, i.e., the W_{ij} terms in eq. (7)?

These quantities are calculated from Green's functions as shown in eqs. (13) and (15). Thus, the accuracy in calculating the relative electrostatic energy of a protonation state is determined by the accuracy in calculating $\Phi(\alpha, \beta)$. For $\alpha = \beta$, we apply the surface integral method to obtain the reaction field potential. For $\alpha \neq \beta$, we use the electrostatic potential as determined from the finest grid containing both charges.

The error in the calculated electrostatic potential is examined for a charge of $1e$ located: (1) at the center of the sphere, and (2) 1 Å below the surface of the sphere. The potentials calculated for these cases are determined by the finite difference method and compared to the analytical solution for charges within a sphere.[2]

ERRORS IN $\Phi(\alpha, \alpha)$

The reaction field potential, i.e., the potential at the protein charge that results from the polarization charge and the ionic atmosphere, is important for determining intrinsic pK_as, because it reflects the degree of solvation of the charged and uncharged amino acid. As described above, the reaction field potential can be calculated in two ways: the subtraction method and the boundary integral method. We compare the results for these two methods with the analytical solution for the two cases in Table I. Both methods for calculating the reaction field potential give values that are within 1% of the analytical solution. This is true for both positions of the charge; however, a 1% error in the reaction field potential for the charge located 1 Å below the molecular surface results in a larger absolute error. Because the self-energy of the charge is given by $W = q\Phi_{RF}/2$, the error in the self-energy of a point charge located 1 Å below the molecular surface will be $\sim 0.1\ pK_a$ units.

It is interesting that both finite difference methods give similar results. The systematic error is likely the result of how the dielectric interface is represented on the finite difference grid. Although the dielectric interface is known analytically, it is represented by a grid, which introduces discretization error into the dielectric interface.

TABLE I.
Reaction Field Potential for a Charge of 1e Located at Two Positions in the r = 30 Å Sphere.

System	Method	Φ_{RF} (pK/e)
Charge at center	Kirkwood (analytic)	−1.982
	Finite difference (subtraction method)	−1.987
	Finite difference (surface integral method)	−1.991
Charge 1 Å below surface	Kirkwood (analytic)	−28.28
	Finite difference (subtraction method)	−28.47
	Finite difference (surface integral method)	−28.52

ERRORS IN $\Phi(\alpha, \beta)$

Because the $\Phi(\alpha, \beta)$ terms are present in both eqs. (13) and (15), errors in $\Phi(\alpha, \beta)$ will affect the calculation of both intrinsic pK_as and interaction energies between charged sites. Atoms with partial charge are distributed throughout the interior of a protein; thus, the accuracy, $\Phi(\alpha, \beta)$, is set by how accurately the potential resulting from a charge can be computed throughout the interior of the protein. We can use the two spherical test systems to compare the potential generated by the charge at points inside the $r = 30$ Å sphere and compare this to the analytical result.

Following the analysis of Gilson et al.,[23] we calculated the electrostatic potential from a source charge at points located on the surfaces of spheres of different radii that are centered on the source charge. Each successive sphere has a radius $r = n$ Å, where $n = 1, 2, \ldots$. Each sphere was evenly covered with 500 points, and only points on the sphere that were deeper than 1 Å from the molecular surface were considered (because we are concerned with potentials at the positions of atom centers, all points at which the potential is to be calculated are at least an atomic van der Waals radius from the molecular surface). The electrostatic potential was calculated at all internal points for each sphere.

The average and maximum errors in the calculated potential for the two positions of the source charge are shown in Figures 3 and 4. For both cases, the calculated potential deviates little from the analytic value.

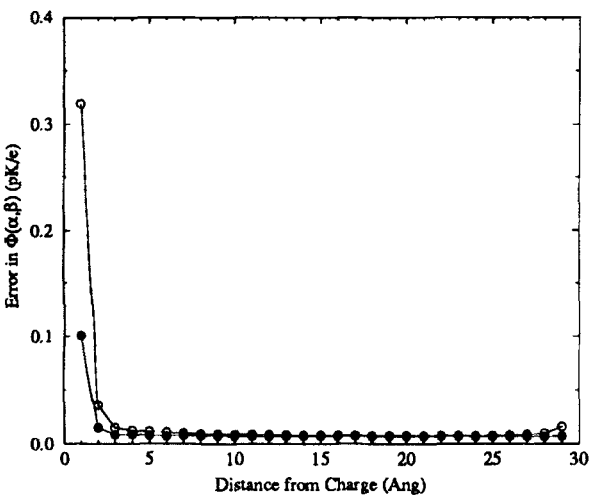

FIGURE 3. Average and maximum error in the calculated electrostatic potential caused by a 1e charge positioned at center of a 30-Å sphere.

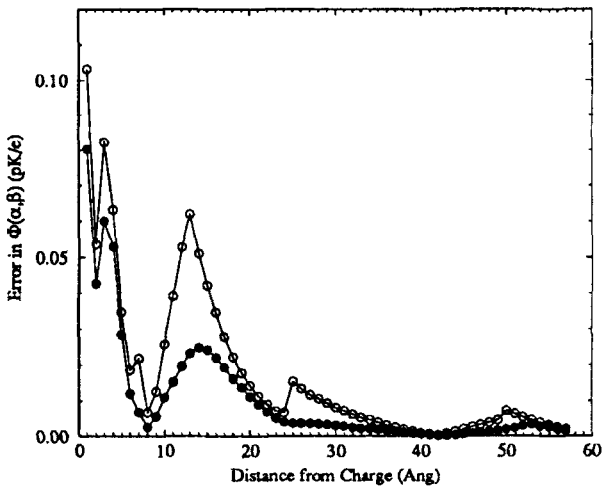

FIGURE 4. Average and maximum error in the calculated electrostatic potential caused by a 1e charge positioned 1 Å below surface of 30-Å sphere.

For the case with the charge positioned at the center of the low dielectric sphere (Fig. 3), the error is significant only very close (~ 2 Å) to the source charge. These errors will have little effect on calculated pK_as. Only bonded atoms approach these distances (and, in the continuum model, the interaction between bonded atom does contribute to the solvation energy). However, the interaction is also proportional to the product of the partial charges on the atoms, which will be less than $1e$ for bonded atoms. Therefore, ~ 0.3 pK_a units represents an upper bound on the error resulting from the $\Phi(\alpha, \beta)$ terms in eqs. (13) and (15).

The case in which the charge is positioned 1 Å below the molecular surface is probably more relevant because titrating sites tend to be accessible to solvent in most proteins. In this case, the error in the electrostatic potential very close to the charge is small. However, deviations further from the charge are more significant. The error in the calculated potential has an oscillation that peaks near the grid boundary for each successive grid in the focusing procedure. For a particular grid, the error in the calculated potential starts to increase as the source charge is approached, but then drops as the distance becomes small enough to be within the next grid in the focusing procedure. At this point, the finer grid gives more accurate potentials than the coarse grid, resulting in an abrupt drop in the error. It is difficult to assess the cumulative impact of these errors in $\Phi(\alpha, \beta)$ on the calculated intrinsic pK_as and site–site interactions, but it appears to be small.

CONVERGENCE

Although the accuracy in the calculated $\Phi(\alpha, \beta)$ for $\alpha \neq \beta$ is limited by the resolution of the finest grid that contains the two points, the calculated reaction field potential [i.e., $\Phi(\alpha, \beta)$ for $\alpha = \beta$] can be shown to converge as the finite difference grid is refined by focusing. The values of Φ_{RF} at each focusing refinement for the two test cases are shown in Figures 5 and 6.

For both systems, the subtraction method and the surface integral method converge. However, both differ from the analytic solution. The error is systematic and remains constant after the finite difference grid lies completely within the low dielectric, indicating that it is caused by the discretization of the dielectric interface. Once the finite difference grid lies entirely within the low dielectric region, further focusing of the grid does not improve the accuracy.

Also interesting is the increase in error for the surface integral method at very high levels of focusing. The subtraction method appears to be more stable. This probably results from the fact that, once inside the low dielectric region, each successive focusing iteration contributes an error associated with the interpolation from the coarse grid needed to define the boundary potential on the fine grid. In the subtraction method, these errors are common to both grids, and therefore, subtract out, whereas no such cancellation occurs in the surface integral method.

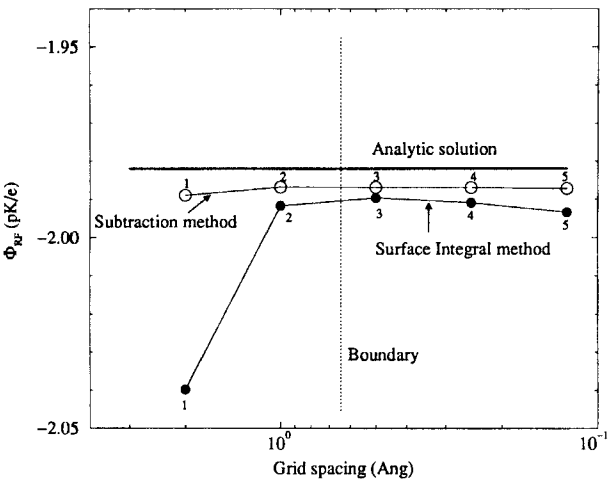

FIGURE 5. Reaction field potential for a 1e charge at the center of a 30-Å sphere as a function of grid spacing. For each level of focusing, we reduced the grid spacing by a factor of 1 / 2; the focusing iteration number is shown for each data point. Both the subtraction method (open circles) and the boundary integral method (filled circles) compare well with the analytical result (horizontal line). The maximum grid spacing that permits the grid to lie entirely within the low dielectric region is labeled "Boundary."

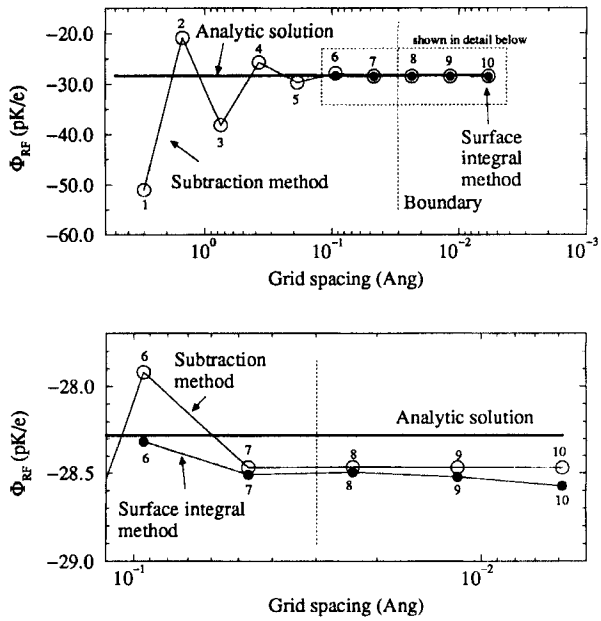

FIGURE 6. Reaction field potential for a 1e charge at a point 1 Å below the surface of a 30-Å sphere as a function of grid spacing. For each level of focusing, we reduced the grid spacing by a factor of 1 / 2; the focusing iteration number is shown for each data point. The lower graph is a blow-up of the upper one. Both the subtraction method (open circles) and the boundary integral method (filled circle) compare well with the analytical result (horizontal line). The maximum grid spacing that permits the grid to lie entirely within the low dielectric region is labeled "Boundary."

Sensitivity Analysis of Continuum Model

We now turn to the calculation of pK_as from the continuum electrostatic model and finite difference method described above. We calculated pK_as for titrating sites in the photosynthetic reaction center (RC) from *Rb. sphaeroides*. For several of these residues, calculated pK_as differed with those inferred from experimental measurements (see Beroza et al.[12] for a full discussion). Our purpose here is to assess the sensitivity of the calculated pK_as to the uncertainties in the parameters of the model and to determine whether such sensitivity might be responsible for the disagreement between theory and experiment.

CALCULATION OF pK_aS

We used the reaction center coordinates as deposited in the Brookhaven Protein Data Bank (prerelease structure P4RCR, date 09-SEP-91).[27] The coordinates of the heavy atoms (i.e., nonhydrogen atoms) were taken from the Protein Data Bank. Hydrogen atoms were added using the computer program InsightII (Biosym Technologies, San Diego, CA). Where possible, positions of polar hydrogens were adjusted by hand to optimize hydrogen bonding. Further details of the RC calculation are presented elsewhere.[12]

Intrinsic pK_as and site–site interactions were calculated from Green's functions for atoms in the 155 titrating sites in the reaction center that were resolved in the X-ray structure. The titration curves for these residues were determined by Monte Carlo[12] sampling of the protonation states of the protein as pH was varied from 3 to 12 in 1.0 pH increments. The pK_a for a residue was estimated by linear interpolation.

For each titrating residue in the protein, a Green's function [i.e., $\Phi(\alpha, \beta)$ in eqs. (13) and (15)] must be obtained for each atom whose charge changes when the amino acid changes protonation state. This requires a number of separated finite difference calculations, each of which may take ~ 8 focusing iterations to converge (see Fig. 6). This can be quite time consuming. Although the most accurate results were obtained using a focusing factor of 0.5 and grid dimensions of ± 33, we have found that accurate results can be obtained more quickly by reducing the grid dimensions to ± 21 (i.e., 42 grid points across the entire grid) and reducing the focusing factor to 1/4. The resulting discrepancy from the more accurate values was < 0.1 pK units for the intrinsic pK_as of all titrating sites. For these calculations, focusing is halted once the finite difference grid lies entirely within the low dielectric region. The parameters for the continuum electrostatic model and finite difference technique are shown in Table II.

For the purpose of the sensitivity analysis, we assume that there is some uncertainty in the parameters of the continuum model. The parameters involved are: (1) the atomic van der Waals radii and solvent probe radius, which together determine the dielectric interface; (2) the protein (internal) and solvent (external) dielectric constants; and (3) the partial charges that represent the polar bonds and the charged titrating sites in the protein.

TABLE II.
Parameters for the Continuum Electrostatic Model and Finite Difference Method.

Parameter	Value
Solvent probe radius	1.4 Å
Ion exclusion length	2.0 Å
Ionic strength (KCl)	50 m*M*
Internal dielectric constant	4.0
External dielectric constant	80.0
Focusing factor	0.25
Convergence criterion	10^{-5}

To estimate the sensitivity of the results to the parameters of the dielectric model, pK_as were calculated with various values for these parameters. Only residues that had a calculated pK_a above 3 and below 12 were included in the statistics. If the pK_a shift was so large that the residue no longer titrated in the pH range 3 to 12, it was not included in the statistics (this occurred for only two or three residues). As a result average pK_a shifts are slightly underestimated. The sensitivity of the calculated pK_a to parameters of the model are summarized in Table III and discussed below for each parameter.

Discussion of the Parameters

van der WAALS RADII AND SOLVENT PROBE RADIUS

The region of space that is considered to be internal to the protein (and assigned a low dielectric constant) is determined by the molecular surface. The shape of this surface depends on the van der Waals radii of the protein atoms and the size of the sphere that represents a solvent molecule. The atomic van der Waals radii used in the calculations (C: 1.9 Å; N: 1.65 Å; O: 1.6 Å; H: 1.0 Å; S: 1.9 Å) were taken from Yang et al.[9] and are similar to those of the PARSE parameter set,[28] which has been optimized to reproduce small molecule solvation energies from a continuum electrostatic model. Atomic partial charges were taken from DISCOVER.[29] To assess the sensitivity of calculated pK_as to variations in the molecular surface, we calculated the pK_as for the reaction center for two cases: (i) all atomic radii increased by 10%, and (ii) all atomic radii decreased by 10%. In addition, we examined the sensitivity to size of the solvent probe, by varying its radius by ± 0.1 Å.

TABLE III.
Sensitivity of pK_a to Variation of Parameters.

Variation [a]	Ave. (SD)[b]	Buried residues[c]	Surface residues[d]
R_{vdW} + 10%	0.4 (0.7)	1.1 (1.1)	0.2 (0.3)
R_{vdW} − 10%	0.5 (0.7)	1.3 (1.0)	0.3 (0.3)
R_{solv} = 1.3 Å	0.2 (0.5)	0.6 (0.9)	0.1 (0.1)
R_{solv} = 1.5 Å	0.2 (0.4)	0.5 (0.6)	0.1 (0.1)
ϵ_p = 2	1.9 (2.1)	3.9 (3.1)	1.3 (1.2)
ϵ_p = 3	0.9 (1.2)	2.1 (2.0)	0.5 (0.5)
ϵ_p = 5	0.5 (0.8)	1.2 (1.3)	0.3 (0.3)
ϵ_p = 6	0.8 (1.0)	1.7 (1.6)	0.5 (0.4)
Point charge	1.7 (2.1)	3.6 (3.3)	1.1 (1.1)
Dipoles +10%	0.2 (0.3)	0.4 (0.6)	0.1 (0.1)
Dipoles −10%	0.2 (0.3)	0.4 (0.5)	0.1 (0.1)

[a] R_{vdW}: atomic van der Waals radii (varied by ±10%); R_{solv}: solvent probe radius (standard value = 1.4 Å); ϵ_p: protein dielectric constant; point charge: titrating site represented as a point charge; dipoles +10%: permanent dipoles on nontitrating sites increased by 10%; dipoles −10%: permanent dipoles on nontitrating sites decreased by 10%.
[b] Average (ave.) and standard deviation (SD) for the shift in pK_a for all titrating sites that have an average protonation between 0.2 and 0.8 somewhere in the pH range 3 to 12 using the standard parameters.
[c] Averages of all titrating sites inaccessible to a 3-Å probe sphere.
[d] Averages of all titrating sites accessible to a 3-Å probe sphere.

The sensitivity of the calculated pK_as to these parameters is shown in rows one through four in Table III. The pK_as are more sensitive to changes in the van der Waals radii than changes in the solvent probe radius. This can be explained by the details of the geometry of the molecular surface. The closest approach of the molecular surface to an atomic nucleus (which is where the charges are centered in the model) is the convex part of the surface, which is defined by the van der Waals radius of the atom. Changes in the van der Waals radius alter this distance directly, while changes in the solvent probe radius only affect the area of the convex surface that is exposed, not its distance from the nucleus.

The sensitivity is four to five times greater for internal residues than for external residues. Residues that border the molecular surface of the protein have a solvent accessibility that is very similar to their corresponding model compound (i.e., the isolated amino acid in solution). For these residues, the effects of an altered molecular surface are the same for the model compound and for the amino acid bound to the protein. This is not the case for buried residues. Changes in the atomic and solvent radii will affect the model compound energy a great deal, but the changes in the protein's molecular surface will be farther from the charge.

DIELECTRIC CONSTANT, $\epsilon(\mathbf{r})$

The value of the dielectric constant at a point in the protein–solvent system is determined by the solvent accessibility of that point. If solvent molecules can make contact with the point, it is part of the high dielectric volume, if not, it belongs to the low dielectric volume. Although the dielectric constant of water is known, the correct choice for the protein dielectric constant in not clear. The dielectric constant of liquid acetamide[13] (which, like a polypeptide, has polar carbonyl and amide groups) and computational studies on globular proteins[30] suggest a value of four, although the applicability of this choice for short-range electrostatic interactions in proteins has not been demonstrated.

Changes in the external dielectric constant (ϵ_w) had little effect on calculated pK_as (data not shown), while changes in the internal dielectric constant (ϵ_p) had a large effect (see rows 5 to 8 in Table III). Like the van der Waals and probe radii, changes in the internal dielectric constant had the largest effect on the pK_as of residues that are poorly solvated, because they have both larger solvation energies and stronger interactions with other charges.

CHARGE DISTRIBUTION, $\rho(\mathbf{r})$

The distribution of free charge, $\rho(\mathbf{r})$, in the protein was represented by partial charges located at the atomic nuclei (charges were taken from the DISCOVER force field,[29] from Biosym Technologies). Unlike fixed charges and permanent dipoles, titrating charges change value during titration. Atoms in titrating residues will have two charge values, one for the unprotonated form of the amino acid, and one for the protonated form.

Two aspects of the charge distribution were varied to determine their effect on amino acid titration. First, to assess the sensitivity of the results to the charge distribution on a titrating site, the partial charges from the force field were replaced with a single point charge on an atom on the amino acid side chain (C_γ for aspartic acid, C_δ for glutamic acid, C_ζ for arginine, N_ζ for lysine, and C_ϵ for histidine). The effect of this much simpler charge model for the titrating sites is shown in row 9 of Table III. Large changes in

calculated pK_as result, especially for buried residues. Second, to assess the sensitivity of the calculated pK_as to the charge distribution in the nontitrating part of the protein, charges that represent the permanent dipole moments of polar bonds were varied by $\pm 10\%$ (rows 10 and 11 in Table III). Overall, the effects on pK_as were small (< 0.5), although they were larger for buried residues than for surface residues.

Conclusions

We have presented a finite difference method for calculating amino acid pK_as in a protein from a continuum electrostatic model. The computational method was based on the Green's function for a point charge in the protein (i.e., the solution to the Poisson–Boltzmann equation with a unit point charge as the source). The relative energies of the different protonation states were obtained from the set of Green's functions for the atoms in the titratable amino acids. Titration curves were then determined from statistical mechanics.

The Green's function formulation for the electrostatic energy of a protonation state breaks the problem up into pieces that can be calculated with high accuracy. Each Green's function is calculated from a series of grids that are centered on the generating charge, which permits focusing to arbitrarily small grid spacing. Convergence in the reaction field potential can be demonstrated, but only after many focusing iterations. For small meshes, the reaction field potential can be obtained directly from the finite difference grid by using a boundary integral method. This avoids the need to perform a separate finite difference calculation, a method commonly used to subtract off the Coulomb singularity at the generating charge. The accuracy of the calculated reaction field potential for both methods is similar and is limited by the finest grid that contains part of the molecular surface. Finer meshes do not improve the accuracy. For spherical test systems, the potential calculated at points away from the generating charge agrees well with analytical solutions. Based on the discrepancies with the analytical result, we conclude that, within the context of the model, it is possible to obtain a precision in calculated pK_a values of a few tenths of a pK_a unit.

Having established the accuracy of the numerical method, we performed an analysis of the sensitivity of calculated pK_a values to variations in the parameters of the continuum electrostatic model. A general trend that emerged in the sensitivity analysis was that the calculated pK_as for poorly solvated amino acids were more affected by changes in the parameters than those calculated for well-solvated amino acids.

It is not straightforward to compare the sensitivity of the results to parameter variation, because the uncertainty of each parameter is unknown. However, based on our calculations, it appears that calculated pK_as are most sensitive to the representation of the charge distribution on the titrating site and the choice of internal dielectric constant, followed by the choice of atomic van der Waals radii. Variation in the strength of the dipole moments of polar bonds and the radius of the solvent probe have the smallest effect on calculated pK_as.

From the analysis, it seems clear that the sensitivity of the calculation to parameter values is much greater than the error resulting from the numerical solution to the electrostatic energy, at least for spherical test systems for which analytical solutions are available. However, in the case of the photosynthetic reaction center, this sensitivity is probably not large enough to account for the large discrepancies between calculated and experimentally measured pK_as (see Beroza et al.[12] for discussion).

Acknowledgments

We thank Donald Bashford, George Feher, and Melvin Okamura for helpful discussions. This work was supported by the National Science Foundation (MCB-89-15631 and UCB-DMS-9213353), Office of Naval Research (N00014-923-1263), and the National Institutes of Health (GM13191 and GM08326).

References

1. K. Linderstrom-Lang, *C. R. Trav. Lab Carlsberg*, **15,** 1 (1924).
2. C. Tanford and J. G. Kirkwood, *J. Am. Chem. Soc.*, **79,** 5333 (1957).
3. J. Warwicker and H. C. Watson, *J. Mol. Biol.*, **157,** 671 (1982).
4. T. You and S. Harvey, *J. Comput. Chem.*, **14,** 484 (1993).
5. R. Zauhar and R. Morgan, *J. Mol. Biol.*, **186,** 815 (1985).
6. D. Bashford and M. Karplus, *Biochemistry* **29,** 10219 (1990).
7. P. Beroza, D. R. Fredkin, M. Y. Okamura, and G. Feher, *Proc. Natl. Acad. Sci. USA*, **88,** 5804 (1991).
8. D. Bashford and K. Gewert, *J. Mol. Biol.*, **224,** 473 (1992).

9. A.-S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig, *Proteins*, **15**, 252 (1993).
10. H. Oberoi and N. M. Allewell, *Biophys. J.*, **65**, 48 (1993).
11. J. Antosiewicz, J. A. McCammon, and M. K. Gilson, *J. Mol. Biol.*, **238**, 415 (1994).
12. P. Beroza, D. R. Fredkin, M. Y. Okamura, and G. Feher, *Biophys. J.*, **68**, 2233 (1995).
13. C. Tanford and R. Roxby, *Biochemistry*, **11**, 2192 (1972).
14. D. Bashford and M. Karplus, *J. Phys. Chem.*, **95**, 9556 (1991).
15. M. K. Gilson, *Proteins*, **15**, 266 (1993).
16. J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, New York, 1975.
17. M. E. Davis and J. A. McCammon, *Chem. Rev.*, **90**, 509 (1990).
18. K. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301 (1990).
19. M. K. Gilson and B. H. Honig, *Proteins*, **3**, 32 (1988).
20. N. Rogers and M. Sternberg, *J. Mol. Biol.*, **174**, 527 (1984).
21. F. M. Richards, *Annu. Rev. Biophys. Bioeng.*, **6**, 151 (1977).
22. M. L. Connolly, *J. Appl. Cryst.*, **16**, 548 (1983).
23. M. K. Gilson, K. A. Sharp, and B. H. Honig, *J. Comp. Chem.*, **9**, 327 (1988).
24. C. J. F. Böttcher, *The Theory of Dielectric Polarization*, vol. 1, Elsevier, Amsterdam, 1973.
25. V. Mohan, M. E. Davis, J. A. McCammon, and B. M. Pettitt, *J. Phys. Chem.*, **96**, 6428 (1992).
26. B. Luty, M. E. Davis, and J. A. McCammon, *J. Comp. Chem.*, **13**, 768 (1992).
27. F. C. Bernstein et al., *J. Mol. Biol.*, **112**, 535 (1977).
28. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.*, **98**, 1978 (1994).
29. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
30. M. K. Gilson and B. H. Honig, *Biopolymers*, **25**, 2097 (1986).